

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	Sig Harold Badt, Jr.	Docket:	139161
Serial No.:	101676,590	Art Unit:	2609
Filed:	October 1, 2003	Examiner:	Josiah J. Hernandez
Title:	Multi-Modal Input Form with Dictionary and Grammar		

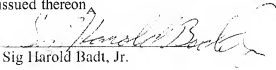
DECLARATION UNDER 37 C.F.R. § 1.131

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

I, Sig Harold Badt, Jr., declare as follows:

1. This declaration is to establish completion of the invention of the above-identified application, Serial No. 101676,590, in the United States at a date prior to September 30, 2003.
2. It is believed that the effective date of the Examiner-cited reference, U.S. Patent Application Publication No. 200510071171, entitled "Method and System for Unified Speech and Graphic User Interfaces," is its filing date of September 30, 2003.
3. At least as early as March 27, 2003, I conceived the invention described and claimed in the above-identified application, as evidenced by the Invention Disclosure form dated March 27, 2003 (copy attached to this Declaration as Exhibit "A") that I submitted to Alcatel, my employer and the assignee of this patent application.
4. It can be appreciated that the invention was made in a commercial environment and that it normally takes the company some finite amount of time to act on a submitted Invention Disclosure and to then file a patent application therefore. In due course thereafter, an outside law firm was contacted to prepare the patent application. I worked diligently with the attorney preparing the patent application from that time until the application was filed on October 1, 2003.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

By: 
Sig Harold Badt, Jr.

Date: 5/10/03

EXHIBIT "A" TO RULE 131 DECLARATION
ATTORNEY-CLIENT PRIVILEGED CONFIDENTIAL INFORMATION

Local Docket No. _____

Alcatel Reference No. ____



ALCATEL USA INVENTION DISCLOSURE FORM

Please e-mail a **soft copy** of this Form to Jerri Pearson at jerri.pearson@usa.alcatel.com and send a **signed paper copy** to Jerri (972 477-9128, Alcanet 2867-9128) at M/S LEGL2. This Form is available on the Alcatel USA Intranet Legal Department site.

Invention Title: Multimodal Input Form With Dictionary and Grammar

Inventors:

Full Name Sig Badt Jr.	Employee No. 18239	M/S CTO2	Phone 972-477-1093	Alcanet
Business Division CTO	Alcatel Company	Citizenship U.S.A.	E-mail Address hal.badt@alcatel.com	
Supervisor Name, M/S, Phone No. Dieter Kopp, +49 7 11 821-32145				
Home Address 302 Trailridge	City, State, Zip Code Richardson, Texas 75081			County U.S.A

Full Name	Employee No.	M/S	Phone	Alcanet
Business Division	Alcatel Company	Citizenship	E-mail Address	
Supervisor Name, M/S, Phone No.				
Home Address	City, State, Zip Code			County

Full Name	Employee No.	M/S	Phone	Alcanet
Business Division	Alcatel Company	Citizenship	E-mail Address	
Supervisor Name, M/S, Phone No.				
Home Address	City, State, Zip Code			County

I, one of the above listed inventor(s), hereby assign and transfer my entire interest in the invention described in this Invention Disclosure Form and the full and exclusive right to any patent therefor to Alcatel USA Sourcing, L.P., having its principal place of business at 1000 Coit Road, M/S LEGL2, Plano, Texas 75075, for the sum of one dollar and/or other good and valuable consideration, the receipt of which is hereby acknowledged.

Inventor Signature(s): 1) _____ Date: _____
 2) _____ Date: _____
 3) _____ Date: _____

Witness Signatures: I have read and understand this invention disclosure:
 1) _____ Date: _____

This invention disclosure is a privileged Attorney/Client communication provided to Alcatel legal counsel for the purpose of obtaining legal advice on the described invention.

ATTORNEY-CLIENT PRIVILEGED

CONFIDENTIAL INFORMATION

Local Docket No. _____

Alcatel Reference No. _____

2) _____ Date: _____

Local Docket No. _____

Alcatel Reference No. _____

**FIT (Fiche D'Information Technique)
TECHNICAL INFORMATION SHEET
Alcatel USA Invention Disclosure Form**

Title: Multimodal Input Form With Dictionary and GrammarAuthor(s) of this FIT: Sig Badt Jr.Date: March 27, 2003Originating Business Division/Unit: CTOOther Affected Business Divisions: MSD and FSD**1. What is the technical problem that was to be solved?**

It is possible for a human user to input information into a computer using a voice dialog. Because of limitations in some current automatic speech recognition software, the human user may only be allowed to say certain things at certain points in the dialog. The problem is that the human user may not know what he or she is allowed to say.

2. What were the best existing solutions (known to the inventor)?

The system can be designed in such a way that is obvious to most human users what should be said at every point in the human/computer dialog. The system designer may try to consider all possible things a human user might want to say at any point in the dialog. The human user may be trained in the use of the system.

3. Why were these existing solutions not good enough?

All of the above solutions may fail. There may be no obvious thing to say. There may be so many possible things the human user may say that the system designer cannot explicitly list them all. Many users of the system may have no access to training.

4. What is the basic idea of the new solution described here?

Please make clear how this is different from the existing solutions.

Design a kind of graphical user interface that can be read by the human user. The human user can look at the GUI and know what he or she can or cannot say at any point in the human/computer dialog.

5. Short description of the new solution, including how it accomplishes what it does.

It is usually helpful to give an example and a drawing.

Extra pages or portions of a report may be included.

See Appendix A.

6. Advantages of the new solution compared to the existing ones. Quantify if possible.

With a single specification, the system designer can build a human/computer interface that can be used as a conventional GUI, a combined GUI and voice interface, and a voice-only interface. A human user can use this interface without specific training.

Local Docket No. _____

[REDACTED]

9. [REDACTED]

10. [REDACTED]

11. [REDACTED]

12. [REDACTED]

Local Docket No. _____

Alcatel Reference No. _____

[REDACTED]

13 [REDACTED]

14 [REDACTED]

15 [REDACTED]

16 [REDACTED]

Local Docket No. _____

Alcatel Reference No. _____

Appendix A: Overview of invention

Computer users today are almost all familiar with window-oriented graphical user interfaces called GUIs or point-and-click interfaces. GUIs can be extended to become multimodal interfaces. In multimodal interfaces, the user can input information into the computer by using the mouse and keyboard in the conventional manner or the user can input data with spoken, gestured, or hand written input. The user can also receive graphical or spoken output from the computer.

A software module that makes it possible for a computer to understand spoken input is called an Automatic Speech Recognition or ASR system. With the current state of the art, it is sometimes only possible for an ASR system to recognize a fixed set of few hundred words and phrases at a given time. For example, at a certain moment in a human/computer dialog, it may be possible for the ASR system to recognize the phrase, "Book a flight from Boston to Chicago," but it may not be possible to recognize, "Book a seat from Boston to Chicago." For the purpose of this document we will describe this situation by saying that at a given point in a human/computer dialog the ASR system can only recognize phrases that conform to a limited dictionary and grammar.

The problem to be solved by this invention is that, with voice input, a human user does not always know what is the acceptable dictionary and grammar at the current point in the human/computer dialog. For example, at a given point in a dialog a user may not know if he or she should say "Book a flight" or "Book a seat."

The proposed solution to this problem is to create a form of GUI window that can be read by the user. See Figure 1. By reading the window, the user can know what is the recognizable dictionary and grammar for spoken input at this moment in the dialog. The GUI window can also be used in the conventional point-and-click manner.

In a conventional GUI window there is a bar across the top of the window (sometimes called the grab bar) that contains on its left side the name of the window. In Figure 1, the name of the window is "Book flight." The user reads the window from left to right and top to bottom, so any recognizable spoken input phrase for this window must start with the words "Book Flight". After that must come the word, "from."

In Figure 1, after the word "from", is a GUI object called a pull-down input field. The user may not know what to say when he or she encounters this field. At this point, the user can say the reserved word, "list." See Figure 2. The system responds with a list of all recognizable inputs at this point in the dialog. The user must speak one of the words in this list. If a user encounters a pull-down input field, and the user already knows a recognizable input word, the user can simply say the input word directly.

The user can also use a pull-down input field with a mouse, stylus, or keyboard in the conventional manner.

At the bottom of Figure 1 is a field with the label "leaving at." This is a text-input field. The user may not know what the system can recognize as input to this field. At this point, the user can say the reserved word, "what?" The system responds by speaking the words "time of day." An alternative approach would be to put the words, "time of day" directly above or below the field (not shown in Figure 1). The user can then speak the time of day.

Figure 3 shows a window that allows the user to book more than one flight. There is a bar across the bottom of the window that contains the word "and". When the user reaches the end of the window, the user can either

Local Docket No. _____

Alcatel Reference No. _____

say "done" or "and." If the user says "done," the window is complete. If the user says "and," the window expands as shown in Figure 4. The user can then book a second flight.

Figure 5 shows a window that gives the user the option of opening up an additional window to input more information. At the bottom of Figure 5 is an icon that represents a new window. After the user is finished filling in all the other information in this window, the user can say the words "special request." This causes a new window to appear. See Figure 6. The name of the window is also "special request." The user can input information in the new window. If the user reaches the bottom of the window shown in Figure 5, and has no special requests, the user can say the reserved word "done."

In Figure 5 the new window icon is located at the bottom of the window. This does not have to be the case. The icon can be anywhere in the original window. When the user finishes inputting data in the new window, input focus returns to the original window just after the new window icon.

Punctuation Devices

In order to assist in the human/computer dialog, special signals may have to be passed back and forth between to human and the computer.

Some of these signals indicate that one or the other wants to begin or finish speaking. For example, the human speaker may press and release a designated button to indicate that he or she is about to begin speaking. The speaker may also press the button and hold down the button until he or she is finished speaking. The button may be a physical button or a GUI object. The computer may also display a "microphone open" indication when it can recognize spoken input from the user.

The computer may output a sound of some kind such as a chime or a tone when it is about to begin speaking. The computer may output a second sound when it is finished speaking.

These signals may or may not be necessary depending on the abilities of the ASR system.

The computer may also give a visual indication of the item on the screen that corresponds to the current point in the dialog. The computer may display a moving arrow or the computer may highlight the location on the screen that corresponds to the current point in the dialog.

VOICE-ONLY INTERFACES

The same dictionary and grammar used for a multi-modal GUI interface of the kind described above can also be used for a voice-only interface. A voice-only dialog is the kind that can be conducted over a telephone with no graphic display.

Figure 7 shows a voice-only dialog using the same dictionary and grammar as that used in Figure 1. Figure 7 illustrates a simple voice-only dialog in which the speaker knows the dictionary and grammar that can be recognized by the ASR system before the dialog begins.

Local Docket No. _____

Alcatel Reference No. _____

Figure 8 illustrates a more complex voice-only dialog in which the user knows some, but not all, of the dictionary and grammar that can be recognized by the ASR system before the dialog begins. Again, this is the same dictionary and grammar used in Figure 1. In this example, the computer speaks the "constant text" that the user would have otherwise have read from the screen.

Figure 9 is an example of the use of the reserved word "what" in a voice-only dialog. When the human says "what" the computer replies with the type of input expected next. It does not give an explicit list of all possible inputs.

Figure 10 is an example of the use of the reserved word "list" in a voice-only dialog. The computer replies with an explicit list of all possible inputs at this point in the dialog.

SINGLE DICTIONARY AND GRAMMAR SPECIFICATION

It is possible to build a GUI interface, a GUI plus voice interface, and a voice-only interface of the kinds described above automatically from a single dictionary and grammar specification. A person skilled in the art can design a single formal language that can serve as input to an automatic multimodal interface builder. It is also possible to specify the dictionary and grammar using a drag-and-drop automatic GUI builder similar to the kind commonly used today.

Local Docket No. _____

Alcatel Reference No. _____

Book flight Help

from to

leaving at

Figure 1: The user can read the grammar for the window by looking at the window.

Book flight Help

from to

leaving at

Atlanta
Chicago
Dallas
Denver

Figure 2: The user says, "Book flight from list." The system responds, "Atlanta, Chicago, Dallas, and Denver." This way the user can learn the dictionary for the window.

Local Docket No. _____

Alcatel Reference No. _____

Book flight		Help	
from	Atlanta ▼	to	Chicago ▼
leaving at	10:00 A.M.		
and			

Figure 3: The user says, “Book flight from Dallas to Atlanta at ten A.M. and.” See what happens next in the following figure.

Book flight		Help	
from	Atlanta ▼	to	Chicago ▼
leaving at	10:00 A.M.		
and			
from	Chicago ▼	to	Denver ▼
leaving at	7:00 P.M.		
and			

Figure 4: This figure is a continuation of the previous illustration. After the user says the word “and”, the system responds by extending the window to let the user book a second flight.

Local Docket No. _____

Alcatel Reference No. _____

The screenshot shows a window titled "Book flight" with a "Help" button in the top right corner. Inside the window, there are two dropdown menus for "from" (set to "Atlanta") and "to" (set to "Chicago"). Below these is a text field for "leaving at" containing "10:00 A.M.". At the bottom, there is a label "special request" followed by an empty rectangular input box.

Figure 5: The user says, “Book flight from Atlanta to Chicago at ten A.M. special request.” See what happens next in the following figure.

This figure shows two overlapping windows. The background window is the "Book flight" window from Figure 5. Overlaid on top of it is a new window titled "Special request" with its own "Help" button. Inside the "Special request" window, there is a "meal choice" dropdown menu set to "vegetarian". Below that is a "seat by" section with a dropdown menu currently showing "Isle". Below the "Isle" dropdown are two more options: "window" and "skip", each with a small right-pointing arrow next to it.

Figure 6: This figure is a continuation of the previous illustration. After the user says the words “special request”, the system responds by displaying a new window.

Local Docket No. _____

Alcatel Reference No. _____

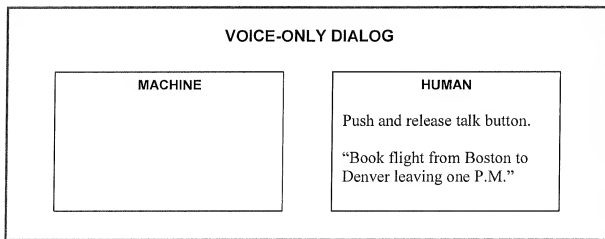


Figure 7: A dialog with the same grammar and dictionary as in Figure 1, but conducted entirely with voice interaction.

Local Docket No. _____

Alcatel Reference No. _____

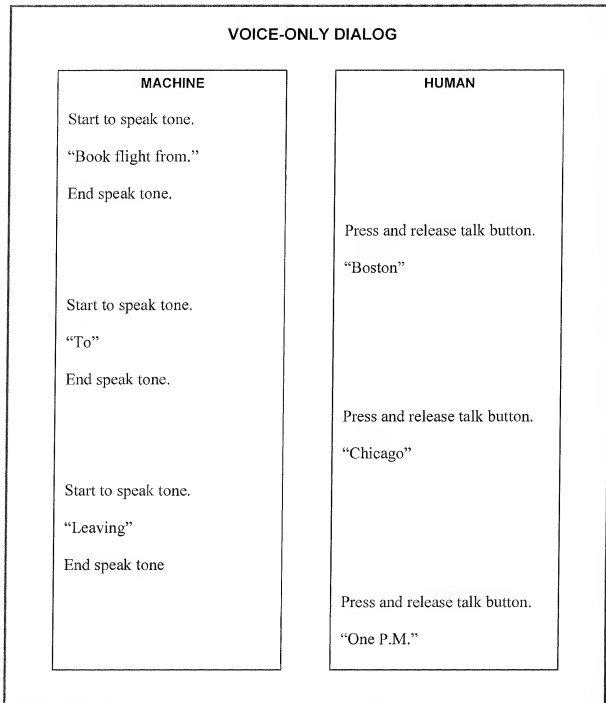


Figure 8: This voice-only dialog uses the same dictionary and grammar as Figure 1 and Figure 7, but with a more interactive conversation. Here the user does not need to know all of the recognizable dictionary and grammar in advance.

Local Docket No. _____

Alcatel Reference No. _____

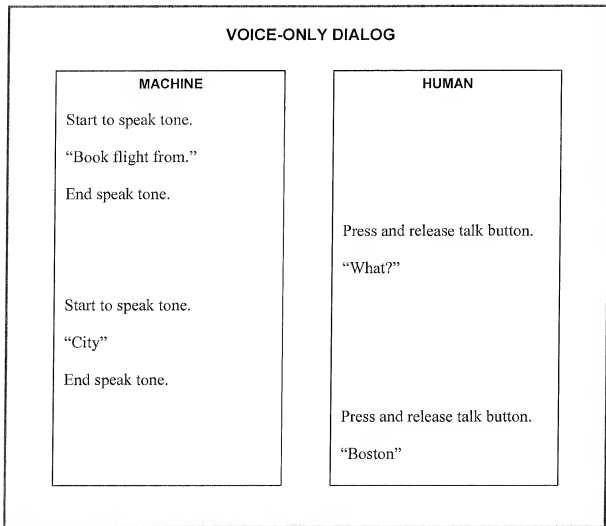


Figure 9: The user speaks the standard reserved word, “What?” The system responds with the type of input expected at this point in the dialog.

Local Docket No. _____

Alcatel Reference No. _____

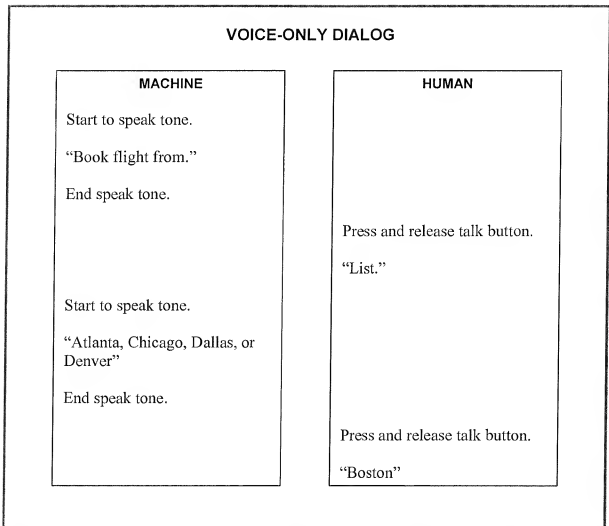


Figure 10: This is an example of the use of the reserved word, "List." The system responds by listing all recognizable responses at this point in the dialog.